
File Organization

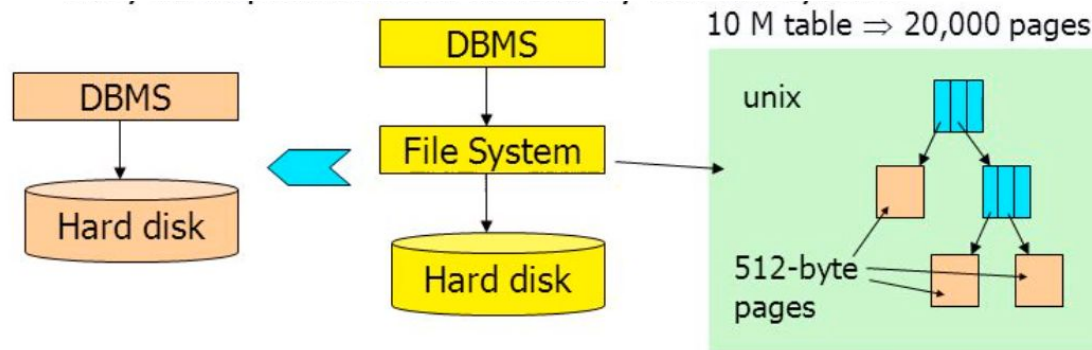
Introducción

Semana 01 / Laboratorio 1.1

Heider Sanchez - hsanchez@utec.edu.pe

Organización de Archivos

- Una base de datos es almacenado como una **colección de archivos**.
- Cada archivo es una secuencia de **registros**.
- Un registro es una secuencia de **campos**.
 - un archivo \Leftrightarrow una tabla
 - un registro \Leftrightarrow una tupla
 - un campo \Leftrightarrow una columna



From CS-HKUST

Conceptos

- **Field:** Esta es la unidad más pequeña de almacenamiento, también conocido como atributo o columna. Un campo tiene dos propiedades: nombre y tipo de dato. El tipo define el tamaño en bytes que requiere su almacenamiento.
- **Record:** Es una colección de campos, también conocido como tupla o fila. Por ejemplo, un registro de empleado puede consistir en los siguientes campos: IdEmpleado, Nombre, Dirección, etc.

Conceptos

- **File:** Es un conjunto relacionado de registros, también conocido como tabla o relación. Un archivo tiene ciertas propiedades como su nombre, tamaño y localización. Los archivos pueden ser archivos de texto o binarios:
 - Los archivos de texto almacenan números como una secuencia de caracteres
 - Los archivos binarios almacenan números en formato binario.
- **File Organization:** Un archivo tiene dos facetas: lógico y físico. El archivo lógico es todo el conjunto de registros. Mientras que el físico muestra como los registros son almacenados físicamente en disco. Organización de Archivos se refiere a la representación física de un archivo.

Conceptos

- **Key:** Este es un atributo que identifica de manera única a cada registro. Similar a la clave primaria de una tabla en una base de datos.
- **Page:** Un archivo es transferido hacia la memoria principal para realizar operación como inserción, modificación eliminación, búsqueda, etc. Si el archivo es demasiado grande, este es fragmentado en páginas de igual tamaño (bloques), el cual es la unidad de intercambio entre el disco y la memoria principal.
- **Index:** Este un puntero a un registro en un archivo, el cual provee acceso eficiente y rápido a los registros.

Operaciones básicas en archivos

- **Read:** Es es el proceso de leer registros desde un archivo, de esta manera se pueden efectuar las operaciones necesarias sobre los campos del registro.
- **Write:** Es el proceso de escribir nuevos registros a un archivo. Pueden ser agregados al final del archivo o insertados en una posición determinada.
- **Delete:** Remueve registros del archivo.

Organización de Registros en un Archivo

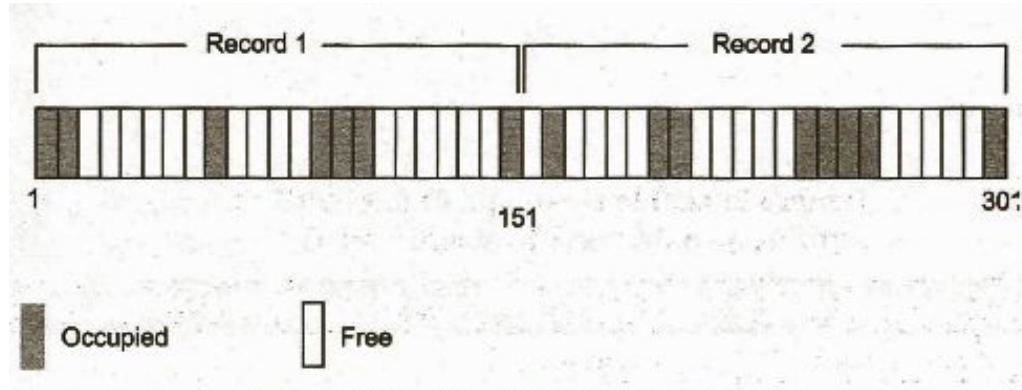
Fixed-Length Records

Todos los registros en un archivo tienen la misma longitud. Cada registro consiste de la misma cantidad de campos y el tamaño de cada campo es fijado por cada registro. Esto asegura fácil ubicación de los valores de cada campo ya que sus posiciones están predeterminadas.

Howard	Paredes	Zegarra	Computacion
Penny	Vargas	Cordero	Industrial
12	12	12	15

Fixed-Length Records

- Desde que cada registro ocupa espacios iguales, identificar el inicio y el fin de cada registro es relativamente simple.



```
n = sizeof(record)
start = n * (i - 1)
end = start + n
```

Fixed-Length Records

- **Problemas:**

- El acceso a los registros es simple pero se corre el riesgo de cruzar bloques.
- La modificación no permite que los registros crucen el límite del bloque.

Fixed-Length Records

- **Alternativas de eliminación del registro i :**

- Mover los registros $i+1, \dots, n$ hacia $i, \dots, n-1$
- Mover el registro n hacia i
- No mover registros, pero enlazar todos los registros liberados en una lista (Free List).

record 0	A-102	Perryridge	400
record 1	A-305	Round Hill	350
record 8	A-218	Perryridge	700
record 3	A-101	Downtown	500
record 4	A-222	Redwood	700
record 5	A-201	Perryridge	900
record 6	A-217	Brighton	750
record 7	A-110	Downtown	600

Alternativa b

Fixed-Length Records

- **Free List:**

- Almacenar la dirección de el primer registro eliminado en el archivo **header**
- Utilice este primer registro para almacenar la dirección del segundo registro eliminado, y así sucesivamente.
- Podemos asumir estas direcciones almacenadas como punteros, ya que "apuntan" a la ubicación de un registro.
- **Optimización:** se puede usar los mismo registros eliminados para guardar los punteros.

header					
record 0	A-102	Perryridge	400		
record 1					
record 2	A-215	Mianus	700		
record 3	A-101	Downtown	500		
record 4					
record 5	A-201	Perryridge	900		
record 6					
record 7	A-110	Downtown	600		
record 8	A-218	Perryridge	700		

Free List

Fixed-Length Records

ENUNCIADO DE LABORATORIO 1.1